



CHAPTER CONTENTS

For each goal, this chapter introduces the objectives for data management to be addressed in the coming decade based upon current knowledge and infrastructure.

Goal 1: Collect and manage data in multiple locations.

Goal 2: Enable users to discover and access data and information via the Internet.

Goal 3: Develop integrated information data products for scientists and decisionmakers.

Goal 4: Preserve data.

Management Structure

National and International Partnerships

One of the goals of the U.S. Climate Change Research Initiative (CCRI) is to enhance and integrate observation, monitoring, and data management systems to support climate process and trend analyses. This chapter lays the strategy for managing integrated data and information for the next decade.

The nature of the concerted effort of the Climate Change Science Program (CCSP) calls for an overarching data policy that provides full and open access to Earth science-related data in a timely fashion. The terms and conditions of exchange and use for this purpose should be agreed to both nationally and internationally. In the early 1990s, the U.S. Global Change Research Program (USGCRP) agreed to data exchange principles that are still adhered to today (see Box 13-1). The governing law for U.S. Government agencies, OMB Circular A130, specifically states that the “open and efficient exchange of scientific and technical government information, subject to applicable national security controls and the proprietary rights of others, fosters excellence in scientific research and effective use of federal research and development funds.” Office of Management and Budget (OMB) Circular A130 establishes agency user charges at the marginal cost of dissemination, including a provision that agencies can plan to “establish user charges at less than cost of dissemination because of a determination that higher charges would constitute a

significant barrier to properly performing the agency’s functions, including reaching members of the public whom the agency has a responsibility to inform”. This lofty standard should be emulated by all participants in the larger endeavor described by this plan.

The need to manage data as a shared national resource in a manner that focuses on the needs of end users has not previously been recognized, nor has the challenge been undertaken in a serious and systematic manner. Climate data are complex and variable as the data are obtained by diverse means, across a broad range of disciplines, for a variety of purposes, and by wide-ranging organizations—individual researchers; institutions; private industry; and federal, state, and local government organizations. These data come in different forms, from a single variable measured at a single point (e.g., a species name) to multi-variate, four-dimensional data sets that may be petabytes (10^{15} bytes) in size.

Although new data sets that integrate information from multiple sources are being developed, current efforts are limited in scope and a significant expansion is required to meet the needs of policymakers and scientists. The challenge is that data are often inconsistently calibrated in space or time—making scientifically sound integration of multiple data sets difficult. No simple data standard can be designed that all data providers will utilize. Moreover, the U.S. government has limited resources to support long-term electronic data management beyond the life of individual investigators’ projects or programs. Currently, no interagency management structure



BOX 13-1

DATA MANAGEMENT FOR GLOBAL CHANGE RESEARCH POLICY STATEMENTS

- The USGCRP requires an early and continuing commitment to the establishment, maintenance, validation, description, accessibility, and distribution of high-quality, long-term data sets.
- Full and open sharing of the full suite of global data sets for all global change researchers is a fundamental objective.
- Preservation of all data needed for long-term global change research is required. For each and every global change data parameter, there should be at least one explicitly designated archive. Procedures and criteria for setting priorities for data acquisition, retention, and purging should be developed by participating agencies, both nationally and internationally. A clearinghouse process should be established to prevent the purging and loss of important data sets.
- Data archives must include easily accessible information about the data holdings, including quality assessments, supporting ancillary information, and guidance and aids for locating and obtaining the data.
- National and international standards should be used to the greatest extent possible for media and for processing and communication of global data sets.
- Data should be provided at the lowest possible cost to global change researchers in the interest of full and open access to data. This cost should, as a first principle, be no more than the marginal cost of filling a specific user request. Agencies should act to streamline administrative arrangements for exchanging data among researchers.
- For those programs in which selected principal investigators have initial periods of exclusive data use, data should be made openly available as soon as they become widely useful. In each case, the funding agency should explicitly define the duration of any exclusive use period.

developed not to fill unmet measurement needs, but instead to improve the quality of existing measurements.

Fulfilling the need for climate and climate-related data that are useful for scientists, planners, and other end users will be a complex task. The overall challenge, then, is:

To provide seamless, platform-independent, timely, and open access to integrated data, products, information, and tools with sufficient accuracy and precision to address climate and associated global changes.

This challenge can be met through development of a system that efficiently links observations to data management and analysis, and ensures timely delivery of climate data and related information and their preservation for future generations (see Figure 13-1). This integration can be implemented using proven and emerging technologies such as the Internet and digital libraries. Specific goals in this effort are:

- 1) Collect and manage data in multiple locations
- 2) Enable users to discover and access data and information via the Internet
- 3) Develop integrated information data products for scientists and decisionmakers
- 4) Identify data to be preserved.

exists to develop and enforce adoption of a complex data management solution. Scientific data that are not institutionally managed are at serious risk of vanishing when the scientist or data collector turns to other projects or retires.

Traditional core activities within data management have been regarded to be data curation—quality control, context-setting (i.e., metadata), preservation, etc.—and distribution of data sets. In order to focus on the needs of the scientists who use the data, we must significantly expand this core to include data discovery (the ability to locate data that are distributed across multiple institutions and disciplines) and data interoperability (changes to how we conduct data management in ways that free users from the productivity losses associated with incompatible formats, unwieldy file sizes, and large non-aggregated collections).

In addition, many of the scientific and decision support needs of the CCSP require analysis and processing of data into specialized products. Even with a large number of measurement systems, there will always be quantities of interest that are either impractical or impossible to measure directly or routinely. Thus, physical models using instrument data as inputs are implemented and can help fill some of the unmet measurement needs of the program. Additionally, products are

These goals will be achieved through implementation of an effective management structure that will ensure interagency coordination of these efforts, scientific and technological guidance, and user input and requirements.

Researchers, planners, and decisionmakers need seamless access not only to information produced by CCSP efforts, but also to the larger scope of information produced by other federal, non-federal, regional, and international programs and activities. These users should be able to focus their attention on the information content of the data, rather than how to discover, access, and use it. The challenge for data management is a system where the user experience will change fundamentally from the current process of locating, downloading, reformatting, and displaying to one of accessing information, browsing, and comparing data with standard tools, such as web browsers, geographic information system (GIS) programs, and scientific visualization/analysis systems, without concern for data format, data location, or data volume.

The strategy for building this framework must be an evolutionary process with a development model based on ongoing interactions with users. In addition, modifications to existing systems and the development of new systems will require use of existing technologies

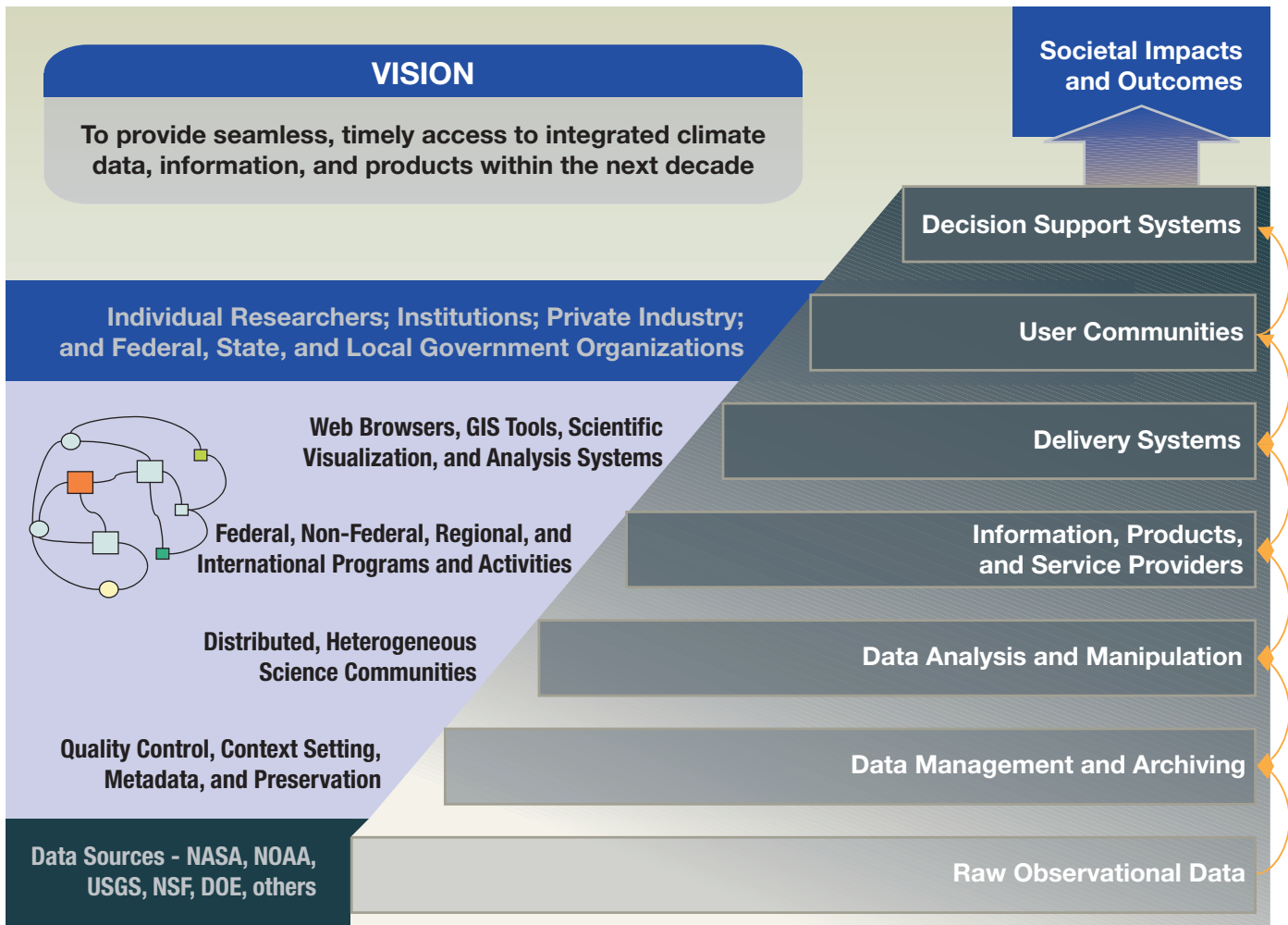


Figure 13-1: Roadmap from data collection to decision support. Source: NASA.

with the vision that the systems would be regularly updated with new technologies to respond to user requirements. Such a framework, with established metadata and quality control/quality assurance standards, mechanisms of transport, protocols, and requirements, will permit data and product providers to contribute their information as well as allow users to query and access the system for relevant information. The challenges to CCSP will be pursuing unprecedented levels of cooperation across current data management institutions and programs and a commitment to mapping the future development and execution of a suitable strategic plan.

The guiding principles for this CCSP data management plan are:

- The measure of success will be the ability of scientists and decisionmakers to access “stand-alone” or “integrated” data and information in a consistent and easily accessed format.
- The value added will be integration—many types of climate data from different suppliers will be available in a manner consistent with user requirements.
- The methods used by data suppliers to deliver data to their “customers” need will evolve with new technology.
- It will be easy for users to discover and access data (local, regional, national, and international).
- The system will be responsive to user feedback.
- The system will preserve irreplaceable data, making use of effective compression technologies where appropriate.

- There will be an open design and open standards process.
- Operations will be reliable, sustained, and efficient.

Goal 1: Collect and manage data in multiple locations.

A distributed system requires CCSP to exploit advances in information technology that enable the development of a distributed data and information system in which data will be collected and managed in multiple locations including federal, state, and local agencies; academic institutions; and non-governmental organizations. Our ability to provide Climate Data Records (CDRs; see Chapter 12, Objective 3.4) and climate information to the community will depend on the interoperability of the system and metadata standards. Long-term archiving and stewardship of the data will be the responsibility of accredited (typically federal) data centers.

Objective 1.1: Develop standard metadata guidelines

Under this objective, CCSP will provide additional specific community-based guidelines for scientific metadata content where and as appropriate. One approach will be to adopt the ISO 19115/TC211 Geographic Information/Geomatics standard (<<http://www.isotc211.org>>), which is built on the Federal Geospatial Data Clearinghouse (FGDC) core standards.



Objective 1.2: Expand collaboration among data providers

CCSP will expand the collaboration between the federal data centers and external (university, commercial, and non-profit) data service providers. This collaboration will build on the strong foundation provided by existing distributed systems, encompassing the data centers established by federal science agencies, such as the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), Department of Energy, U.S. Department of Agriculture, U.S. Geological Survey (USGS), and the National Science Foundation. The data management plan also calls for expanding partnerships with foreign governments, intergovernmental agencies, and international scientific bodies and data networks to provide data that are needed to address the international character of research and decisionmaking. These collaborations should improve access to regional, state, and local data.

Goal 2: Enable users to discover and access data and information via the Internet.

This goal requires a greater emphasis on the development of a framework to respond to the need for integration and communication of information across disciplines and among scientists and policymakers. Multi-agency and multidisciplinary institutional and data resources will need to be targeted to develop standards and processes for sound data management. System upgrades need to include the

implementation of tools to enable communication among multiple data locations. The process of identifying the data requirements of the program on a regular basis, including visualization, analysis, and modeling requirements, needs to be strengthened. Human resources will be required to perform these tasks, particularly individuals with the technical expertise to support user requirements. These needs will be addressed by CCSP (see Figure 13-2).

Objective 2.1: Improve access to data

Several activities are planned under this objective which will enable improved access to data:

- Expand the Global Change Master Directory (GCMD) to facilitate access to data. Agencies will provide descriptions in the format needed for this action.
- Ensure the provision of socioeconomic data collected by federal statistical agencies (e.g., the Census Bureau and the Bureau of Economic Affairs), by resource management agencies (e.g., the U.S. Department of Agriculture, U.S. Army Corps of Engineers, the U.S. Bureau of Reclamation, the U.S. Bureau of Land Management, and the U.S. Fish and Wildlife Service), by energy-related agencies (e.g., the Department of Energy and the Environmental Protection Agency), and by state and local agencies.

Objective 2.2: Management of biological data

Management of biological data will receive priority. Objective 2.3 of the Observing and Monitoring research element focuses on developing new capabilities for ecosystem observations. This is a CCRI priority and is a critical need for evaluating the effects of

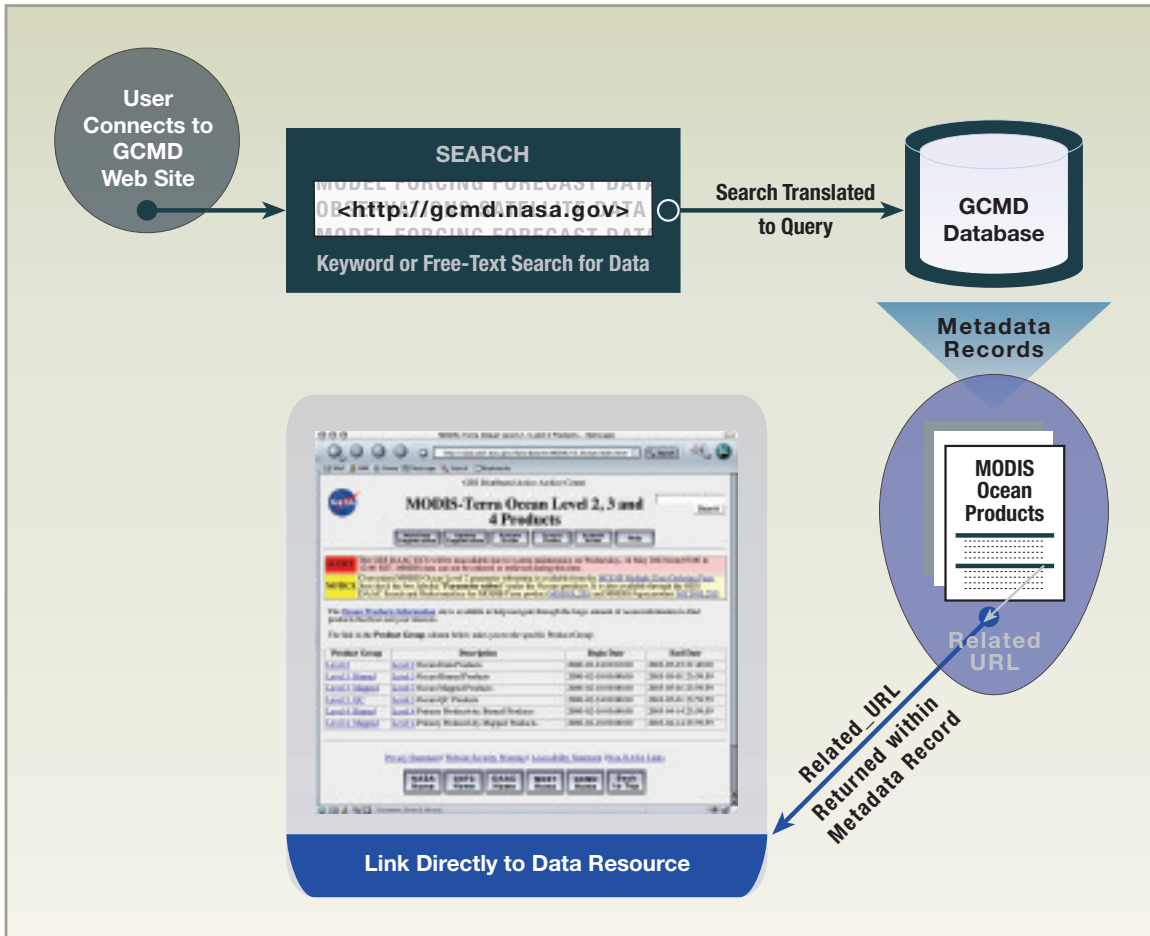


Figure 13-2: Search and direct retrieval of data set information from NASA's Global Change Master Directory. Source: NASA.

climate change on ecosystems. Biological data management is hampered by its requirement for extensive metadata, changes in named taxonomic species, and availability (at present mostly in non-electronic form and in the hands of individual investigators).

Objective 2.3: Data and information portals

Under this objective, CCSP will consolidate agency data information into one portal; that is, an agency home page would provide a mechanism for identifying all available data and information. CCSP will create special, tailored portals for data products of interest to the various CCSP working groups. These portals will use the emerging web metadata clearinghouse technology to allow researchers to locate and access coincident data of interest from various observation systems. This will require implementation of climate-quality data and metadata documentation, standards, and formatting policies that will make possible the combined use of targeted data products taken at different times, by different means, and for different purposes. Additionally, CCSP will work toward supporting the national climate observing system monitoring architecture described in Chapter 12 (Objective 1.7)

Goal 3: Develop integrated information data products for scientists and decisionmakers.

The goal of information analysis and interpretation is to incorporate the multidisciplinary science elements of the CCSP in order to integrate information and provide integrated products. This requires that links between scientists and data managers on one hand and data quality and data products on the other need to be enhanced to provide a more effective translation of user requirements into data products (see Chapter 12, Objective 3.4). Data managers must be able to understand, communicate, and work closely with scientists and others to ensure proper stewardship for the data archive and its distribution. Data managers must be included in scientific working groups and steering committees to guide the integration of data management and science and decision support. CCSP will ensure data quality and preservation by making data management an integral part of any observing or data collection program. Decision support needs will set the priorities for integrated products and help to define and address data management issues associated with the integrated products.

Objective 3.1: Establish links between data providers and decisionmakers

A dialogue needs to be established between data providers and decisionmakers to understand how scientific resources and knowledge are currently used by decisionmakers. Scientific discussion, planning, and implementation are needed for key data issues (e.g., gridding algorithms, gap-filling periods of missing observations) to permit assimilation of many CCSP data products. CCSP data analysis will draw on and promote further advances in data-processing automation, data visualization techniques, and web-based data delivery mechanisms. Activities under this objective include:

- Create a link on the CCSP website where decisionmakers can search for, locate, and link to data and information products identified by CCSP working groups as potentially of significant utility.

- Develop a prototype of the provision of support services for decision support systems. Provide an initial operational capability that interfaces one or more CCSP data systems to one or more decision support systems.
- Implement procedures to solicit climate information requirements from regions, sectors, and users who are using climate projections for management and policy decisions.

Objective 3.2: Application of data products and information

The emphasis on regional data, modeling, and decision support activities is increasing. Researchers and stakeholders are collaborating to develop applications based on research findings. An example of this effort is an experiment to integrate scientists and stakeholders to frame and apply El Niño-Southern Oscillation forecasts and other research products in a variety of regions and economic sectors. Regional environmental data products that can provide up-to-date information on environmental conditions to decisionmakers—and, if appropriate, allow an interactive “If . . . , then . . .” environment—must be anchored in considerations of input and process uncertainties and outcome accuracies. Decision support services must provide information about uncertainty to be of maximum utility to decisionmakers. This objective includes:

- Improve access to climate data and information for addressing regional concerns and issues
- Provide geo-referenced and spatially and temporally averaged socioeconomic data products for integrated studies
- Continually improve and clearly articulate the accuracy of regional data.

Objective 3.3: Harness emerging technologies

CCSP needs to take advantage of emerging information systems such as Digital Libraries (DLs). DL is a paradigm for investment by several agencies that has the potential of becoming the world’s most vital environment for discourse and resources promoting excellence in science and education. Data management could be greatly informed and enabled by the DL technology. The guiding principles for the development of DL are to provide a spectrum of interoperability, to provide one library with many portals, and to leverage the energy and achievements of others. DL’s effort will focus on building a comprehensive library of digital resources and this effort will enhance CCSP’s successful implementation.

Goal 4: Preserve data.

One daunting challenge of the 21st century is the management of the large volume of highly diverse data describing the Earth’s climate. These data are a result of comprehensive observing and monitoring systems and models producing new data sets from the climate observations. The size of the data archives is growing faster than we can derive information from them. For example, NASA’s Earth science data holdings increased by a factor of six between 1994 and 1999; the total amount of data doubled between 1999 and 2000. It is estimated that by 2010, the size of a major U.S. archive for data from NOAA, NASA, and USGS will be 18,000 terabytes (10^{18} bytes). Lessons learned from NASA’s pioneering efforts in handling their current holdings (more than 2,500 terabytes) must be used by



BOX 13-2

EXAMPLES OF DATA PRODUCTS

Chapter 3 –**Atmospheric Composition**

- Improved description of the global distributions of aerosols and their properties.
- A 21st century chemical baseline for the Pacific region against which future changes can be assessed.

Chapter 4 –**Climate Variability and Change**

- New and improved climate data products, including assimilated data from satellite retrievals and other remotely sensed and *in situ* data for model development and testing; consistent and regularly updated reanalysis data sets suitable for climate studies; centuries-long retrospective and projected climate system model data sets; high-resolution data sets for regional studies; and assimilated aerosol, radiation, and cloud microphysical data for areas with high air pollution.
- A paleoclimatic database designed to evaluate the ability of state-of-the-art climate models to simulate observed decadal to century-scale climate change, responses to large changes in climate forcing, and abrupt climate change.

Chapter 5 – Water Cycle

- Integrated long-term global and regional data sets of critical water cycle variables such as evapotranspiration, soil moisture, groundwater, clouds, etc., from satellite and *in situ* observations for monitoring climate trends and early detection of climate change.
- Enhanced data sets for feedback studies, including water cycle variables, aerosols, vegetation, and other related feedback variables, generated from a combination of satellite and ground-based data.

Chapter 6 –**Land-Use/Land-Cover Change**

- Global high-resolution satellite remotely sensed data and land-cover databases.
- Operational global monitoring of land use and land-cover conditions.

Chapter 7 – Carbon Cycle

- Global, synoptic data products from satellite remote sensing documenting changes in terrestrial and marine primary productivity, biomass, vegetation structure, land cover, and atmospheric column CO₂.

- Landscape-scale estimates of carbon stocks in agricultural, forest, and range systems and unmanaged ecosystems from spatially resolved carbon inventory and remote-sensing data.

Chapter 8 – Ecosystems

- Data sets for examining effects of management and policy decisions on a wide range of ecosystems to predict the efficacy and tradeoffs of management strategies at varying scales.
- Synthesis of known effects of increasing CO₂, warming, and other factors (e.g., increasing tropospheric ozone) on terrestrial ecosystems based on multi-factor experiments.

Chapter 9 –**Human Contributions and Responses to Environmental Change**

- Assessments of the potential economic impacts of climate change on the producers and consumers of food and fiber products.
- Elevation maps depicting areas vulnerable to sea-level rise.

the community. In addition, new technologies need to be developed that will enable us to keep all data needed for long-term global change research, reducing the need to prioritize which data will be archived. This endeavor would also consider lessons learned from communities that already handle this volume of data (e.g., defense intelligence, commercial video streaming).

Objective 4.1: Enhance the data management infrastructure

Telecommunications bandwidth capacity must be adequate to accommodate the movement of these larger data volumes as they progress through an information cycle including measurements, distributed scientific analyses, science models, predictions, decision support tools, assessments, and policy and management decisions. Increased levels of bandwidth will become available through government research, development, and funding; commercial availability and acquisition; and non-profit sector partnering. It is important to keep in mind that the evolutionary realization of this vitally needed infrastructure must be continually planned. Another critical area requiring enhancement is the

development of new technologies for storage of large volumes of data and information.

Objective 4.2: Preserve historical records

At the same time, many important heritage data sets face a growing risk of loss due to deterioration of paper records, obsolescence of electronic media and associated hardware and software, and the gradual loss of experienced personnel (see, e.g., Figure 13-3). We look to these historical records, from which we can derive long-term trends, to help provide the missing pieces of the overall climate puzzle. The primary focus under this objective will be to identify and rescue data that are at risk of being lost due to media deterioration, poor accessibility, or limited distribution.

Objective 4.3: Support an open data policy

Another data management challenge is data policy—described as the set of rules, regulations, laws, or agreements governing the access and use of data. Database protection legislation, enacted in Europe and proposed in the United States, has raised concerns that the flow of scientific information may become much more constrained.

Many of these policies are in conflict with each other and the challenge will be to understand these conflicts and chart a course that benefits all. This will necessitate the close interaction of and negotiation between the database rights holders and users, in order to strike a balance between protection and fair use (NRC, 1999f). Compiling long-term climate-quality data sets from which long-term climate trends can be derived will be greatly affected by the future data policies of national and international bodies. CCSP will develop and implement guidelines for when and under what conditions data will be made available to users other than those who collected them.

Management Structure

Working in partnership with members and representatives of the research community in federal agencies and academia, and with appropriate committees of the National Research Council, CCSP will seek to identify the data requirements of the program on a regular basis, including visualization, analysis, and modeling requirements. Priority attention will be given to those observations and data that are central to a specific research element but for which requirements are not currently being met, or that exist but are not part of a publicly available data system. Accomplishment of these goals will require an integrated management structure that involves the CCSP agencies with oversight by members of the external community. A Data Management Steering Committee composed of federal, state, academic, and industry managers and decisionmakers will provide oversight, priorities, coordination, and recommendations to the CCSP Data Management Working Group (DMWG). DMWG will be responsible for the preparation, implementation, and periodic review of data management activities, and publication of annual reports describing milestones achieved and future activities. Close links via shared membership will be maintained with the Observing and Monitoring boards and councils as described in Chapter 12.

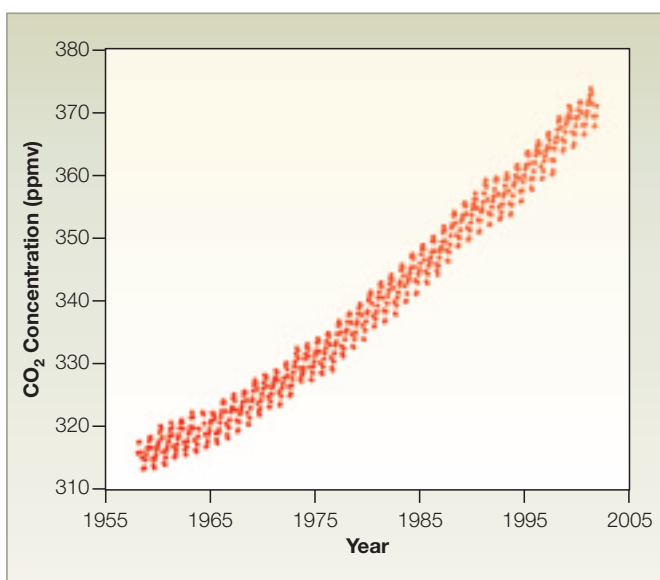


Figure 13-3: Changes in atmospheric carbon dioxide (CO₂) concentration at Mauna Loa, Hawaii, over time. This figure illustrates the critical need to preserve historical data. Source: Dave Keeling and Tim Whorf, Scripps Institution of Oceanography.

National and International Partnerships

CCSP will facilitate access to the data and information required and generated as part of this program. A critical need for observations and data are identified throughout this plan. Box 13-2 provides examples from each research chapter that illustrates the type of data products to be generated; Box 13-3 provides examples of the type of information products. Note that this latter box does not reflect output from the various modeling activities described in each chapter.

BOX 13-3

EXAMPLES OF INFORMATION PRODUCTS

Chapter 3 – Atmospheric Composition

- A *State of the Atmosphere 2006* report that describes and interprets the annual status of the characteristics and trends associated with atmospheric composition, ozone layer depletion, temperature, rainfall, and ecosystem exposure.

Chapter 4 – Climate Variability and Change

- Documented impacts of climate extremes on regions and sectors, and evaluations (both positive and negative) of the implications should climate change in the future.

Chapter 5 – Water Cycle

- Assessment reports on the status and trends of water flows, water uses, and storage changes for use in analyses of water availability.

Chapter 6 – Land-Use/Land-Cover Change

- Report on the regional and national impacts of different scenarios of land use and land cover on water quality and quantity.

Chapter 7 – Carbon Cycle

- *State of the Carbon Cycle* report focused on North America.

Chapter 8 – Ecosystems

- Reports describing the potential consequences of global and climatic changes on selected arctic, alpine, wetland, riverine, and estuarine ecosystems; selected forest and rangeland ecosystems; selected desert ecosystems; and the Great Lakes.

Chapter 9 – Human Contributions and Responses to Environmental Change

- Assessments of the potential health effects of combined exposures to climatic and other environmental factors (e.g., air pollution).



The generation of U.S. and global data products will require cooperation with national and international data centers and institutions. The CCSP will utilize and participate in the development of the data discovery and data interoperability framework being advanced by other programs such as the U.S. Integrated Ocean Observing System effort. CCSP will coordinate its activities with international programs to take advantage of emerging data management and information tools and technologies and sharing of climate change data and information. Examples of international programs that actively engage in data management include the World Data Center system, which functions under the guidance of the International Council of Scientific Unions (ICSU) and facilitates international exchange of scientific data; the World Climate Research Programme, which sponsors multiple major projects involving international cooperation and data collection with guidance by a Joint Scientific Committee; the International Human Dimensions Programme on Global Environmental Change Data and Information System (IHDP/DIS), which links social science data centers and scientists; and the Data Management Coordination Group of the Joint World Meteorological Organization/

Intergovernmental Oceanographic Commission Technical Commission for Oceanography and Marine Meteorology, which is currently developing an Oceans Information Technology Pilot Project.

CHAPTER 13 AUTHORS

Lead Authors

Margarita Conkright, NOAA and CCSP
 Wanda Ferrell, DOE
 Clifford Jacobs, NSF
 Martha Maiden, NASA

Contributors

Vanessa Griffin, NASA
 Steve Hankin, NOAA
 Thomas Karl, NOAA
 Chet Koblinsky, NASA